

SOUND SOURCE LOCALIZATION SYSTEM, AND SOUND REFLECTING
ELEMENT

FIELD OF THE INVENTION

The present invention relates to a sound source localization system, a sound source localization method, a sound reflecting element useful for the sound source localization system, and a method for forming the sound reflecting element. It more particularly relates to a high precision sound source localization system, a sound source localization method, a sound reflecting element useful for the sound source localization system, and a method for forming the sound reflecting element, in which the sound source position including the elevation data can be acquired with high precision even if the system comprises a smaller number of microphones.

BACKGROUND OF THE INVENTION

Conventionally, to enhance the sound source localization performance with a microphone array, a processing system capable of making the simultaneous input for multiple channels comprising a number of microphones has been needed. This processing system allows a driving member to be

1 controlled efficiently to face a sound source position.
2 However, if a number of microphones are arranged to acquire
3 the sound source position, there is an inconvenience that
4 the total cost of the system is increased. Therefore, an
5 attempt for reducing the number of microphones has been
6 made. However, in the conventional attempt for reducing the
7 number of microphones, if the number of microphones was
8 reduced, there was an inconvenience that the information for
9 giving a full directionality toward the sound source was
10 lacked. Also, employing the conventional method, there was
11 an inconvenience that the localization of the sound source
12 was more likely to be affected by the surrounding noise, a
13 variation in the property of sound source and the transfer
14 characteristics of the room, although the sound source
15 position was acquired to some extent under the conditions
16 where the properties of the sound source were specified and
17 the measurement environment was managed.

18 In the estimation of the sound source position employing a
19 small number of microphones, various methods have been
20 hitherto proposed. For example, a binaural hearing method
21 employing two microphones has been well known. This method
22 involves using a head transfer function (HRTF), measuring
23 the head transfer function at a binaural position, disposing
24 a sound source for generating a reference sound at various
25 azimuths, ranges and elevations, and adding the transfer

1 characteristics at the binaural position to acquire the
2 positional information. The above head transfer function is
3 obtained by deciding experimentally the transfer
4 characteristics from the sound source to the ears, including
5 the influences of the head, chest, and concha, for each
6 model, but has a disadvantage of having poor universality.

7 Moreover, the localization of the sound source employing the
8 above head transfer function is made by measuring the
9 signals from the sound source, and selecting a signal
10 consistent with an acoustic spectrum given by the head
11 transfer function measured in advance to acquire the sound
12 source position. Accordingly, the method employing the head
13 transfer function allows the localization of the sound
14 source more or less correctly in principle, if the sound
15 source is a reference sound source. However, since the
16 acquisition of sound source position employing the head
17 transfer function makes the use of a dip or a peak arising
18 in the head transfer function as a characteristic key
19 profile, the sound source position may be possibly
20 misjudged, when the sound source has the dip or peak.
21 Therefore, in the present state of affairs, the acquisition
22 of sound source position employing the head transfer
23 function is employed more frequently in the sound
24 reproduction than the acquisition of sound source position.

1 More particularly, the conventional method for acquiring the
2 sound source position was disclosed in Okuno et al., "Are a
3 pair of ears sufficient for robot audition?", The journal of
4 The Acoustical Society of Japan, vol. 58, no. 3, pages
5 205-210, in 2002, in which the acquisition of sound source
6 position employing two microphones was examined. With this
7 method, the range and azimuth are acquired, employing the
8 ILD (Interaural Level Differences) and the ITD (Interaural
9 Time Difference) obtained from the head transfer function.
10 In the above acquisition of sound source position employing
11 two microphones, the azimuth and range of the sound source
12 can be acquired by measuring the above characteristic values
13 from the acoustic spectrum observed. However, only with
14 these bits of information, the range may not be acquired
15 when the sound source for the acoustic spectrum is located
16 in direct front.

17 The reason is that in, the interaural level differences and
18 the interaural time difference are constant, even when the
19 range is different. Also, the sound source localization
20 method employing the interaural level differences and the
21 interaural time difference are not effective for vertical
22 localization. The reason is that as long as the azimuth
23 and range are common, the interaural time difference and the
24 interaural level differences are common, even if the
25 elevation varies. From the above reason, to acquire the

1 sound source position including the range and elevation, it
2 is considered that there is a need for taking cues on the
3 reverberation the ~~deformation~~ information of the acoustic
4 spectrum, like the monaural hearing as will be described
5 later, and also pointed out that there is a need for further
6 examination.

7 Apart from the binaural hearing, an attempt for acquiring
8 the sound source position by a method of what is called the
9 monaural hearing has been made. The monaural hearing for
10 localization of the sound source is similar to the manner
11 that the man acquires the range to the sound source, in
12 which a larger sound with less reverberation is perceived as
13 the near sound, and a smaller sound with more reverberation
14 is perceived as the distant sound. Employing the loudness
15 of sound and the reverberation as described above, the range
16 to the sound source position is roughly acquired. However,
17 the loudness of sound depends on the sound source of object,
18 and the level of reverberation depends on the experimental
19 environment of acoustic spectrum as well. In the case of
20 man, the information about the sound source of object and
21 the environment, including the visual information, may be
22 compensated by performing a high level information
23 processing, and utilized to acquire the range to the sound
24 source. This processing is practically difficult to
25 implement on a signal processing system comprising an

1 information processing apparatus only based on a pure
2 routine process.

3 Also, in the review for the method for human to acquire the
4 sound source position, it has been found that the azimuth
5 and elevation to the sound source attenuates the spectrum in
6 a specific frequency range under the influence of the head
7 and concha. However, the acquisition method is affected by
8 the properties of the sound source for the same reason as
9 explained for the method employing the head transfer
10 function, and is difficult to implement.

11 Regarding the use of a reflecting plate similar to the
12 concha, a parabolic reflector for collecting a remote subtle
13 sound has been offered by positively utilizing its
14 reflection characteristics. Figure 15 shows a schematic
15 constitution of the parabolic reflector that has been
16 offered. The parabolic reflector 100 as shown in Figure 15
17 comprises a reflecting plate 102 for reflecting a sound wave
18 101 from a distant sound source and a microphone 104 for
19 collecting the reflected sound wave. The reflecting plate
20 102 is roughly formed from a paraboloid, and the microphone
21 104 is disposed at a focal point position of the paraboloid.
22 The sound wave 106 reflected from the reflecting plate 102
23 is focused at the focal point to efficiently collect the
24 sound, but there is no function of acquiring the sound

1 source position.

2 Moreover, in an apparatus such as a robot or a sound
3 handling KIOSK terminal that is an object spoken to from the
4 man, it is required to make an operation of "facing in that
5 direction", "turning the directivity of a microphone array
6 to the corresponding direction" or "ignoring a distant
7 sound". For this purpose, it is required that the robot or
8 apparatus recognizes the range or direction to the sound
9 source, or the talker, and controls a drive control system
10 to initiate a necessary operation. That is, under the
11 conditions where the kind of signal sound is unknown, there
12 were the disadvantages with the existing technologies that
13 (1) one microphone does not allow the acquisition of sound
14 source position in principle, and (2) the existing system
15 with two microphones does not allow the acquisition of the
16 range in the forward direction and the elevation in the
17 vertical direction.

18 Also, an increased number of microphones are arranged at
19 appropriate positions as conventionally to relieve the above
20 limitations, whereby the acquisition precision is improved.
21 However, due to a packaging constraint of the design cost,
22 it is sought to relieve the above limitations with a smaller
23 number of microphones.

1 As described above, there is a need for a new method and
2 means suitable for acquiring the position of a sound source,
3 employing an information processing system, without the use
4 of the scale of ~~deformation~~ information of spectrum, sound
5 volume or intensity of reverberation needing a high level
6 preliminary knowledge. Also, there is a further demand for
7 a sound source localization system and a sound source
8 localization method in which the range, azimuth and
9 elevation to the sound source are acquired employing the
10 above method and means. Also, there is a further need for a
11 sound reflecting element and a design method for it in which
12 the acquisition of sound source position is excellently
13 made.

14 SUMMARY OF THE INVENTION

15 In light of the above-mentioned problems associated with the
16 prior art, an aspect of the present invention recognizes
17 that the disadvantages of the prior art can be solved as far
18 as the elevation information to a sound source can be
19 analyzed with high precision, employing at least one sound
20 collecting means, more particularly, a microphone, whereby a
21 sound source localization system and a sound source
22 localization method are provided with higher precision.

1 In an example embodiment of the present invention, a sound
2 wave generated from a sound source is reflected inherently
3 according to a sound source position, and recorded as the
4 acoustic data collected with the direct sound. This
5 acoustic data is converted into digital data for later
6 processing and once held in a recording unit. The acoustic
7 data can provide a new cue referred to as a delay
8 ~~deformation~~ information in this invention. Therefore, in
9 this invention, the new scale of "delay ~~deformation~~
10 information" is employed in addition to the conventional
11 cue, without depending on the kind of signal sound source,
12 whereby the disadvantages associated with the prior art in
13 the acquisition of sound source position can be solved.

14 In another aspect, to record acoustic data the present
15 invention provides the delay ~~deformation~~ information with a
16 high inherent property, this invention provides a sound
17 reflecting element for reflecting a sound wave generated
18 from the sound source inherently corresponding to a sound
19 source position to enable the recording, and a processing
20 method for processing the recorded acoustic data.

21 In still another aspect, according to the present
22 invention, there is also provided a sound source
23 localization system comprising a sound reflecting element
24 for generating a delay ~~deformation~~ information corresponding

1 to a relative position between a sound source and sound
2 collecting means, a storage part for storing the acoustic
3 data collected via the sound reflecting element, and a sound
4 source localization part for acquiring a sound source
5 position, employing the acoustic data on which the delay
6 ~~deformation~~ information is superposed. The sound reflecting
7 element of the invention may be formed as a spheroid
8 associated with the relative position between the sound
9 source and sound collecting means to generate the delay
10 ~~deformation~~ information intrinsic to the relative position.
11 The sound source localization part of the invention may
12 comprise a standard template storage part for storing a
13 standard template containing an intrinsic delay ~~deformation~~
14 information generated by a white noise sound source, a
15 background noise template storage part for storing a
16 background noise template, a residual generation part for
17 calculating a residual from the acoustic data, employing the
18 standard template and the background noise template, and a
19 selection part for selecting the standard template giving
20 the least residual, employing the generated residual.

21 In another aspect, according to the invention, there is
22 provided a sound source localization method for acquiring
23 the position of a sound source under the control of an
24 information processing apparatus, the method comprising a
25 step of collecting the acoustic data with a delay

1 ~~deformation~~ information superposed corresponding to a
2 relative position between a sound source and sound
3 collecting means, a step of storing the collected acoustic
4 data in a storage part, and a step of reading the acoustic
5 data with the delay ~~deformation~~ information superposed and
6 acquiring the relative position of the sound source
7 designated by the delay ~~deformation~~ information.

8 BRIEF DESCRIPTION OF THE DRAWINGS

9 The invention and its embodiments will be more fully
10 appreciated by reference to the following detailed
11 description of advantageous and illustrative embodiments in
12 accordance with the present invention when taken in
13 conjunction with the accompanying drawings, in which:

14 Fig. 1 is a view showing the parameters for defining the
15 sound source position and the position in the present
16 invention;

17 Fig. 2 is a view for explaining an essential principle for
18 generating a delay ~~deformation~~ information in this
19 invention;

20 Fig. 3 is a view for explaining an essential principle for

1 forming a reflecting surface of a sound reflecting element
2 in this invention;

3 Fig. 4 is a view schematically showing the reflection of
4 sound wave on the reflecting surface as shown in Fig. 3;

5 Fig. 5 is a view showing the envelope for forming the
6 cross-sectional shape of the sound reflecting element formed
7 in this invention;

8 Fig. 6 is a view showing the sound reflecting elements
9 according to an embodiment of the invention;

10 Fig. 7 is a view showing an arrangement of sound reflecting
11 elements according to the embodiment of the invention;

12 Fig. 8 is a schematic flowchart showing a sound source
13 localization method of the invention;

14 Fig. 9 is a block diagram showing the schematic
15 configuration of a sound source localization system of the
16 invention;

17 Fig. 10 is a block diagram showing the detailed
18 configuration of the sound source localization part of the
19 invention;

1 Fig. 11 is a view showing a standard template and the
2 storage of three-dimensional position coordinates according
3 to the embodiment of the invention;

4 Fig. 12 is a graph showing a delay ~~deformation~~ information
5 obtained in this invention;

6 Fig. 13 is a graph showing the correlation between the delay
7 ~~deformation~~ information generated in the invention and the
8 delay ~~deformation~~ information on design;

9 Fig. 14 is a diagram showing the precision of sound source
10 position acquired in this invention; and

11 Fig. 15 is a view showing the schematic configuration of a
12 conventional parabolic reflector.

13 DESCRIPTION OF SYMBOLS

14 10 ... sound reflecting element
15 12 ... sound collecting means (microphone)
16 14 ... plane
17 16 ... imaginary line
18 18 ... sound reflecting element

1 20 ... talker
2 22 ... sound reflecting element
3 24 ... recording part
4 26 ... sound source localization part
5 28 ... driving element
6 30 ... acoustic data storage part
7 32 ... STP storage part
8 34 ... BNT storage part
9 36 ... PF part
10 38 ... residual storage part
11 40 ... selection part
12 42 ... application execution part

13 **DETAILED DESCRIPTION OF THE INVENTION**

14 The present invention provides methods, systems and
15 apparatus for solving problems associated with the prior
16 art. The present invention recognizes that the
17 disadvantages of the prior art can be solved as far as the
18 elevation information to a sound source can be analyzed with
19 high precision, employing at least one sound collecting
20 means, more particularly, a microphone, whereby a sound
21 source localization system and a sound source localization
22 method are provided with higher precision.

1 In an example embodiment of the present invention, a sound
2 wave generated from a sound source is reflected inherently
3 according to a sound source position, and recorded as the
4 acoustic data collected with the direct sound. This
5 acoustic data is converted into digital data for later
6 processing and once held in a recording unit. The acoustic
7 data can provide a new cue referred to as a delay
8 ~~deformation~~ information in this invention. Therefore, in
9 this invention, the new scale of "delay ~~deformation~~
10 information" is employed in addition to the conventional
11 cue, without depending on the kind of signal sound source,
12 whereby the disadvantages associated with the prior art in
13 the acquisition of sound source position can be solved.

14 To record the acoustic data by providing the delay
15 ~~deformation~~ information with a high inherent property, this
16 invention provides a sound reflecting element for reflecting
17 a sound wave generated from the sound source inherently
18 corresponding to a sound source position to enable the
19 recording, and a processing method for processing the
20 recorded acoustic data.

21 According to the present invention, there is also provided a
22 sound source localization system comprising a sound
23 reflecting element for generating a delay ~~deformation~~

1 information corresponding to a relative position between a
2 sound source and sound collecting means, a storage part for
3 storing the acoustic data collected via the sound reflecting
4 element, and a sound source localization part for acquiring
5 a sound source position, employing the acoustic data on
6 which the delay ~~deformation~~ information is superposed. The
7 sound reflecting element of the invention may be formed as a
8 spheroid associated with the relative position between the
9 sound source and sound collecting means to generate the
10 delay ~~deformation~~ information intrinsic to the relative
11 position. The sound source localization part of the
12 invention may comprise a standard template storage part for
13 storing a standard template containing an intrinsic delay
14 ~~deformation~~ information generated by a white noise sound
15 source, a background noise template storage part for storing
16 a background noise template, a residual generation part for
17 calculating a residual from the acoustic data, employing the
18 standard template and the background noise template, and a
19 selection part for selecting the standard template giving
20 the least residual, employing the generated residual. The
21 standard template storage part of the invention may store
22 the standard template and the sound source position giving
23 the standard template in association. The sound source
24 localization system of the invention may comprise one or
25 more sound reflecting elements, and simultaneously acquires
26 the positional data of the sound source including a range to

1 the sound source, an azimuth and an elevation as the
2 relative position.

3 According to the invention, there is provided a sound source
4 localization method for acquiring the position of a sound
5 source under the control of an information processing
6 apparatus, the method comprising a step of collecting the
7 acoustic data with a delay ~~deformation~~ information
8 superposed corresponding to a relative position between a
9 sound source and sound collecting means, a step of storing
10 the collected acoustic data in a storage part, and a step of
11 reading the acoustic data with the delay ~~deformation~~
12 information superposed and acquiring the relative position
13 of the sound source designated by the delay ~~deformation~~
14 information. The delay ~~deformation~~ information of the
15 invention may be generated by reflection from a spheroid
16 associated with the relative position between the sound
17 source and sound collecting means, and the delay ~~deformation~~
18 information may be generated intrinsic to the relative
19 position. The sound source localization step of the
20 invention may comprise a step of reading out a standard
21 template from a standard template storage part for storing
22 the standard template containing a delay ~~deformation~~
23 information intrinsic to the relative position generated by
24 a white noise sound source, a step of reading out a
25 background noise template from a background noise template

1 storage part for storing the background noise template, a
2 step of calculating a residual from the acoustic data,
3 employing the standard template and the background noise
4 template, and a step of selecting the standard template
5 giving the least residual, employing the generated residual.
6 The selection step of the invention may comprise a step of
7 referring to the selected standard template and acquiring
8 the sound source position corresponding to the standard
9 template. The sound source localization method of this
10 invention may further comprise a step of simultaneously
11 acquiring the range, azimuth and elevation as the relative
12 position from the acquired sound source position to the
13 sound source.

14 According to the invention, there is provided a sound
15 reflecting element for generating ~~a delay deformation~~
16 information corresponding to a relative position between a
17 sound source and sound collecting means, wherein a
18 reflecting surface of the sound reflecting element is
19 designed as an envelope made from a plurality of spheroids
20 that are formed by rotating a plurality of ellipses having
21 the two focal points corresponding to the sound source and
22 the sound collecting mean around an axis connecting the
23 focal points.

24 The plurality of ellipses in this invention may be generated

1 in relation with the elevation between the sound source and
2 the sound collecting means and flatter as the elevation is
3 greater. The reflecting surface in this invention may be
4 designed as an enveloping surface of the plurality of
5 spheroids that are generated by rotating a corresponding
6 ellipse around the axis connecting the focal points.

7 According to the invention, there is provided a formation
8 method of a sound reflecting element for generating a delay
9 ~~deformation~~ information corresponding to a relative position
10 between a sound source and sound collecting means, the
11 method comprising a step of generating a plurality of
12 spheroids by rotating an ellipse having the focal points
13 corresponding to the sound source and the sound collecting
14 mean around an axis connecting the focal points, and a step
15 of forming a reflecting surface by generating an enveloping
16 surface of the plurality of spheroids. The plurality of
17 ellipses in this invention may be generated in relation with
18 the elevation between the sound source and the sound
19 collecting means and flatter as the elevation is greater.

20 A. Constitution of sound reflecting element

21 Figure 1 is a view showing the definition of the range,
22 azimuth and elevation for use in the present invention. In
23 Figure 1, the microphones M1 and M2 as sound collecting

1 means are employed, in which the azimuth, range and
2 elevation are represented as the position coordinates
3 measured from a middle point between the microphones M1 and
4 M2. A sound source SS is separated away by a predetermined
5 range r from the middle point between the microphones. In
6 the above coordinates, the sound source position can be
7 represented in the Cartesian coordinate system (x, y, z) or
8 polar coordinate system (r, θ, ϕ) in this invention. In the
9 following, the acquisition of elevation is explained as a
10 specific embodiment in this invention, but the invention is
11 applicable to the acquisition of any sound source position
12 collected in the scale of angle and range, in addition to
13 the azimuth and elevation.

14 This invention essentially involves a path difference
15 between the sound wave directly collected from the sound
16 source and the reflected wave reflected from a reflecting
17 surface of the sound reflecting element, such that the shape
18 of sound reflecting element is configured to relate the
19 position of sound source with the path difference. In the
20 invention, the sound reflecting element is configured
21 essentially as a set of elliptic curves. Conventionally,
22 for an elliptic curved surface, it is well known that the
23 sound wave produced from one focal point of the ellipse is
24 reflected to the other focal point. Figure 2 shows the
25 typical properties of the ellipse. As shown in Figure 2,

1 the cross section of the reflecting surface is configured
2 using the ellipse in which the sound source is disposed at
3 one focal point A and the microphone is disposed at the
4 other focal point B in this invention. In an arrangement as
5 shown in Figure 2, a sound wave S_r starting from the focal
6 point A is collected at the same focal point B, even if
7 reflected at any position on the wall. Employing the
8 ellipse as the reflecting surface, it follows that the
9 reflected wave always has a certain path difference $(2a-f)$
10 as defined by the elliptic curve from a sound wave S_d not
11 reflected and directly going from the focal point A to the
12 focal point B.

13 Taking notice of the path difference, it was reviewed to
14 positively utilize the path difference for the localization
15 of the sound source in this invention. Herein, considering
16 an application mode of the realistic sound reflecting
17 element in the acquisition of sound source position, it is
18 important in the realistic configuration that the microphone
19 is fixed relative to the sound reflecting element, and the
20 sound source such as a talker is moved. Thus, the
21 properties of the reflecting surface are examined, when the
22 position of the microphone is fixed at one focal point B,
23 and the position of the focal point A is changed to have the
24 position of the sound source at the other focal point A. In
25 Figure 3, the maximum range for judging the position of the

1 sound source is defined, and the noise is judged as beyond
2 the maximum range. In Figure 3, the sound source is moved
3 from the supposed farthest position f_{\max} to the supposed
4 nearest position f_0 . At the same time, the shape R of an
5 envelope for the ellipses with the focal points f_{\max} and f_0 is
6 shown when the sound source is moved from the supposed
7 farthest position f_{\max} to the supposed nearest position f_0 in
8 Figure 3. As shown in Figure 3, when the focal point A
9 (sound source position) is closer to the microphone, the
10 ellipse has a rounded shape similar to the circle, or when
11 the focal point A (sound source position) is far away from
12 the microphone, the ellipse has a collapsed shape. Also, as
13 the focal point A is farther, the left end shape
14 approximates asymptotically the parabola. In this
15 invention, the shape of sound reflecting element is
16 essentially configured as the envelope of elliptic curves
17 that are formed in connection with the movement of sound
18 source position.

19 Figure 4 is a view schematically showing the reflection of
20 sound wave from the sound source position A, when the
21 reflecting surface is configured as the shape of envelope as
22 shown in Figure 3. As shown in Figure 4, when the sound
23 wave from the nearer sound source position is reflected at a
24 rear portion of the elliptic curve, its reflected wave is
25 collected at the focal point B that is the microphone

1 position. On the other hand, when reflected near an end
2 portion of the elliptic curve, the sound wave is diffused
3 because the angle is not consistent. Therefore, a major
4 portion of the reflected wave detected is occupied by the
5 wave reflected at the rear portion of the sound reflecting
6 element. Similarly, for another sound source position, it
7 has been found that the reflection position to make a major
8 reflected wave component in accordance with its sound source
9 position is generated when the reflecting surface R of sound
10 reflecting element is configured from the envelope. That
11 is, in this invention, it has been found that the major
12 reflected wave intrinsic to the sound source position is
13 generated when the sound reflecting element is formed with
14 the reflecting surface containing the enveloping surface of
15 ellipses. On the other hand, a path difference between the
16 major reflected wave and the direct wave is accompanied with
17 a delay time, which is equivalent to the path difference as
18 defined by the corresponding ellipse.

19 Moreover, the present inventors have reviewed the elevation
20 determination when the envelope of ellipses is employed as
21 the reflecting surface. Figure 5 shows an envelope of
22 elliptic curves and a shape RS of sound reflecting element
23 corresponding to the envelope when the range between the
24 microphone position B and the sound source position A is set
25 at the designed value, and the elevation θ is changed from

1 the supposed lowest angle θ_0 to the supposed highest angle
2 θ_{\max} . As explained in Figure 4, if the sound reflecting
3 element RS is formed by the envelope, the sound wave from
4 the sound source at low angle has its major reflected wave
5 reflected at the bottom portion of the sound reflecting
6 element, while the sound wave from the sound source at high
7 angle has its major reflected wave reflected at the top
8 portion of the sound reflecting element. This major
9 reflected wave is accompanied with a delay time
10 corresponding to the path difference defined by the
11 corresponding ellipse. That is, the reflected wave
12 intrinsically corresponds to the sound source position.

13 Though this invention has been described above in detail in
14 connection with the cross-sectional shape of the reflecting
15 surface, the shape of the sound reflecting element of the
16 invention is required to be provided in the three dimensions
17 in reality. In this invention, the three-dimensional shape
18 of the reflecting surface of the sound reflecting element
19 for reflecting the sound wave can be formed as the
20 enveloping surface of a plurality of spheroids produced by
21 rotating the corresponding ellipse around an axis connecting
22 the focal point on the side where the microphone is placed
23 and the focal point where the sound source position is
24 located.

1 Figure 6 shows a specific embodiment of the sound reflecting
2 element that is configured according to the invention. For
3 the sound reflecting element 10 of the invention as shown in
4 Figure 6, the tangential line with each spheroid
5 corresponding to the sound source position is also shown to
6 easily recognize the shape. As shown in Figure 6, the sound
7 reflecting element 10 of this invention is configured by
8 cutting the enveloping surface of the spheroid into a size
9 easily employed. Figure 6A is a perspective view of the
10 sound reflecting element 10 as seen from the side of a
11 concave face, and Figure 6B is a perspective view of the
12 sound reflecting element 10 as seen from the side of a
13 convex portion. As shown in Figure 6, the sound reflecting
14 element 10 of the invention has a bottom portion 10a
15 composed of an ellipsoid having a large eccentricity and an
16 upper end portion 10b composed of an ellipsoid having an
17 increased eccentricity, and is narrowed toward the upper end
18 portion 10b in accordance with the elevation.

19 In the sound reflecting element 10 of the invention, the
20 microphone 12 is disposed at one common focal point of the
21 spheroid making up the sound reflecting element 10. Also,
22 the microphone 12 is disposed at a position symmetrical to
23 the sound reflecting element 10 on a plane 14 containing the
24 bottom portion 10a. In the embodiment as shown in Figure 6,

1 the position of the microphone 12 is located on the side of
2 the sound reflecting element 10 above an imaginary line 16
3 connecting the transverse ends of the sound reflecting
4 element 10. However, it may take any position as far as the
5 reflected wave from the sound reflecting element 10 is
6 received uniformly with the noise suppressed in this
7 invention. Also, the sound reflecting elements 10 of the
8 invention may be connected vertically with the plane 14 as
9 the boundary.

10 Figure 7 is a perspective view showing an arrangement of the
11 sound reflecting element 10 according to the embodiment of
12 the invention. In the arrangement as shown in Figure 7, the
13 sound reflecting elements 10 and 18 are disposed as one
14 pair. The sound reflecting elements 10 and 18 have the
15 microphones 12 and 12a disposed in the same configuration as
16 shown in Figure 6. Moreover, in the arrangement of the
17 sound reflecting element as shown in Figure 7, the sound
18 reflecting elements 10 and 18 are faced in the same
19 direction and suitable for acquiring the sound source
20 position in the direction where the concave portions of the
21 sound reflecting elements 10 and 18 are opposed. The sound
22 reflecting element of the invention can essentially acquire
23 the elevation of the sound source position, employing one
24 sound reflecting element, but employing the sound reflecting
25 elements as one pair as shown in Figure 7, the range,

1 elevation and azimuth to the sound source position may be
2 decided simultaneously.

3 Also, if the overall shape of the sound reflecting element
4 is designed to be small, the path difference between the
5 direct wave and the major reflected wave is shortened. To
6 observe its influence precisely, a high sampling frequency
7 is required. In the specific embodiment of the invention,
8 when the elevation to the sound source is 0° and 72° , and if
9 the path difference between the direct wave and the major
10 reflected wave is about 9.5 cm, a delay time difference of
11 about 0.28 ms is produced. When the sampling frequency is
12 48 KHz, this delay time is equivalent to a difference of
13 about thirteen samples. That is, theoretically, it follows
14 that the elevation to the sound source has a maximum
15 resolution of 13 levels to discriminate the elevation from 0°
16 to 72° . In this invention, if the overall shape is designed
17 to be half in size while keeping the resolution, it is
18 required that the sampling frequency is doubled to 96 KHz.
19 Also, if the overall size of the shape is designed to be
20 double, the same resolution is attained even when the
21 sampling frequency is halved or 24 KHz.

22 B. Sound source localization method and system of the
23 invention

1 Figure 8 is a schematic flowchart of a sound source
2 localization method according to the invention. In the
3 sound source localization method of the invention as shown
4 in Figure 9, the acquisition of elevation is made employing
5 the sound reflecting element as explained in the section A.
6 In the sound source localization method of the invention as
7 shown in Figure 8, at step S10, the acoustic data such as
8 voice data is collected via the sound reflecting element
9 from the microphone, converted into digital data, employing
10 an AD converter and stored in memory. At step S12, an
11 observed profile is calculated from the acoustic data in
12 accordance with a method as disclosed in detail in "Speech
13 Enhancement by profile fitting method", O. Ichikawa et al.,
14 IEICE Transactions on information and System, VoL. E86-D,
15 No. 3, pp. 514-521, Mar. 2003, and at the same time, a
16 standard template (STP) and a background noise template
17 (BNT) that are stored in respective storage parts are read
18 out. At step S14, a residual $\Phi_{n,\omega}$ between the observed
19 profile and a linear combination of the standard template
20 and the background noise template is calculated, and stored
21 in an appropriate memory.

22 At step S16, it is determined whether or not there is left
23 any standard template to be further read out. In this
24 manner, the residuals are calculated for all the standard
25 templates. Then, at step S18, the residual $\Phi_{n,\omega}$ is

1 normalized for each subband frequency, and stored in memory.
2 At step S20, the minimum value of the normalized residuals
3 $\Phi_{n,\omega}$ is decided. Then, at step S22, the sound source position
4 corresponding to the standard template giving the minimum
5 value of the calculated residuals is acquired, and selected
6 as the sound source position. At step S24, the coordinates
7 of the sound source position registered corresponding to the
8 selected sound source position are output in an appropriate
9 format to the driving element for controlling the acquired
10 sound source position.

11 As the method for calculating the residual in this
12 invention, a profile fitting method (hereinafter referred to
13 as a PF method) is applied. Particularly in the preferred
14 embodiment of the invention, the PF method is desirably
15 employed. The PF method is a noise suppression method as
16 disclosed in "Speech Enhancement by profile fitting method",
17 O. Ichikawa et al., IEICE Transactions on information and
18 System, VoL. E86-D, No. 3, pp. 514-521, Mar. 2003, whereby
19 the noise is removed, employing the observed profile from
20 the sound source where the elevation, azimuth and range are
21 defined. However, the PF method is also appropriately
22 employed for a process for estimating the sound source
23 position in this invention.

24 The observed profile for use in a process of the specific

1 embodiment of the invention means a power distribution at
2 each subband frequency that is observed by processing an
3 audio signal recorded by the microphone with a delay sum
4 array, and allocating the angle of directivity of the delay
5 sum array from the maximum value to the minimum value. In
6 this invention, the standard template means a template
7 profile normalized in the area from a two-dimensional
8 observed profile including the delay ~~deformation~~ information
9 recorded via the sound reflecting element employed in the
10 invention and measured in advance for a white noise sound
11 source at the known position in which the direction of
12 allocating the angle of directivity is taken along the axis
13 of abscissas and the power is taken along the axis of
14 ordinates.

15 Also, the background noise template in this invention means
16 a template profile normalized in the area from an acoustic
17 profile observed by placing a white noise sound source at
18 the noise sound source position, in which the width of
19 allocating the angle of directivity is given according to
20 the number of sampling channels. In creating the standard
21 template and the background noise template, it is desirable
22 to employ the white noise having a power over the entire
23 frequency band, as previously described. However, the
24 signal and the noise to be actually observed may be employed
25 to approximate the white noise.

1 Moreover, the residual $\Phi_{n,\omega}$ of the invention is given by the
2 following formula.

3 [Formula 1]

4

5
$$\Phi_{n,\omega} = \int_{\min_{\theta}}^{\max_{\theta}} (X_{\omega}(\theta) - \alpha_{n,\omega} \cdot P_{n,\omega} - \beta_{n,\omega} \cdot Q_{\omega}(\theta))^2 d\theta. \quad (1)$$

6 In the above expression, $X_{\omega}(\theta)$ is the power at the subband
7 frequency ω in which the audio signal with a delay
8 ~~deformation~~ information superposed through the sound
9 reflecting element of the invention is processed with the
10 angle of directivity of the delay sum array in the θ
11 direction, and here called the observed profile. $P_{n,\omega}(\theta)$ is
12 the template profile stored as the standard template
13 corresponding to the sound source position, and $Q_{\omega}(\theta)$ is the
14 template profile stored as the background noise template.
15 Also, n corresponds to the sound source position.

16 When the PF method is employed for the sound enhancement,
17 the component decomposition should be made for each frame.
18 However, for the sound source localization, the component

1 decomposition should be made once for the average over all
2 the audio frames to allow the acquisition of sound source
3 position. So, $X_{\omega}(\theta)$ may be the average of speaking
4 utterances for several seconds. If $\alpha_{n,\omega}$ and $\beta_{n,\omega}$ are decided
5 using the above formula, the residual $\Phi_{n,\omega}$ is obtained.
6 Moreover, the normalized residual $\bar{\Phi}_{n,\omega}$ is calculated by
7 dividing $\Phi_{n,\omega}$ by the power for each subband and averaging
8 over Ω subbands as defined by the following formula.

9 [Formula 2]

$$10 \quad \bar{\Phi}_n = \frac{1}{\Omega} \sum_{\omega} \frac{\Phi_{n,\omega}}{\int_{\min_{\theta}}^{\max_{\theta}} \{X_{\omega}(\theta)^2\} d\theta} \quad (2)$$

11 Also, the acquisition of sound source candidate position is
12 made by selecting a sample template sound source candidate
13 position \hat{n} so that the normalized residual may be the
14 minimized, and selecting the acquired sound source position,
15 using the following formula (3).

16 [Formula 3]

$$17 \quad \hat{n} = \underset{n}{\operatorname{argmin}} (\bar{\Phi}_n) \quad (3)$$

18 An index of "profile" as used in this invention contains not
19 only the cue of delay ~~deformation~~ information for the

1 acoustic spectrum, but also the cues of the interaural time
2 difference and the interaural level differences as
3 conventionally employed. That is, the method of the
4 invention not only detects the delay ~~deformation~~
5 information, but also makes it possible to employ the cues
6 of the interaural time difference and the interaural level
7 difference as conventionally employed, together with the cue
8 of delay ~~deformation~~ information. Therefore, in this
9 invention, the range, azimuth and elevation required for the
10 acquisition of sound source position can be acquired
11 simultaneously. Accordingly, in the invention, the process
12 for acquiring the sound source position is performed
13 seamlessly, employing a smaller number of microphones than
14 conventionally needed, and the availability of the sound
15 source localization system is expanded. That is, the
16 acquisition of elevation, which was conventionally
17 impossible with the sound source localization method
18 employing as few as one or two microphones, is not dealt
19 with exceptionally, but is processed at the same time with
20 the case of acquiring the angle in the horizontal direction
21 which was conventionally allowed, whereby the process is
22 performed faster. Also, the cue of delay ~~deformation~~
23 information with the sound reflecting element is added to
24 the case for acquiring the angle which was conventionally
25 allowed, whereby the higher precision localization is
26 allowed.

1 Figure 9 is a view showing the schematic configuration of
2 the sound source localization system according to a specific
3 embodiment of the invention. The sound source localization
4 system of this invention comprises a sound reflecting
5 element 22 for collecting and recording voices from the
6 talker 20, a recording part 24 for converting the acoustic
7 data recorded in the sound reflecting element 22 into
8 digital data and storing it, and a sound source localization
9 part 26 for acquiring the sound source position by analyzing
10 the acoustic data. The acquired sound source position
11 information is passed to an application execution part, not
12 shown, in an appropriate format of the coordinates of sound
13 source position such as the Cartesian coordinates (x, y, z)
14 or the polar coordinates (r, θ , ϕ) that is decided employing
15 the registered standard template.

16 The application execution part receives an input of position
17 coordinates and drives the driving element 28 needed in the
18 specific embodiment. The driving element 28 may be a head,
19 a hand, a foot, an eye, a mouth, the body, a leg, or the
20 whole body for the robot, a camera or a microphone for the
21 kiosk apparatus, or a microphone or a camera for a security
22 system. However, the invention is not limited to the above
23 driving elements.

1 Also, the sound source localization system of the invention
2 is implemented as an information processing apparatus
3 roughly comprising a CPU (Central Processing Unit), a
4 memory, an external I/O control device, a modem and an NIC.
5 Moreover, the sound source localization system of the
6 invention is mounted on the apparatus comprising the driving
7 element for the robot being driven by application software,
8 in which a predetermined position of the driving element is
9 controlled and driven by comparing a range difference, an
10 azimuth difference and an elevation difference between the
11 original position and the acquired sound source position.

12 Figure 10 is a detailed functional block diagram showing the
13 functional configuration of a sound source localization part
14 26 included in the sound source localization system of the
15 invention. The sound source localization part 26 shown in
16 Figure 10 is realized by a program executing the sound
17 source localization method that is mounted on the robot,
18 kiosk, cache dispenser, a security device for making an
19 operation by sensing a sound, the program being executed by
20 the CPU to function as each means as mentioned above. As
21 shown in Figure 10, the sound source localization part 26 of
22 the invention comprises an acoustic data storage part 30 for
23 reading out the acoustic data once stored in the recording
24 part as the digital data by the sound reflecting element 22,
25 and storing it for processing, a standard template (STP)

1 storage part 32, and a background noise template (BNT)
2 storage part 34.

3 Moreover, the sound source localization part 26 of the
4 invention comprises a profile fitting (PF) part 36 for
5 calculating the residual, a residual storage part 38 for
6 storing the residual $\Phi_{n,\omega}$ obtained by the PF part 36, a
7 selection part 40 for selecting the standard template giving
8 the minimum residual from the normalized residual, and an
9 application execution part 42 for executing a necessary
10 application.

11 The PF part 36 of the invention reads in the acoustic data,
12 converts it into an observed profile, then reads out the
13 standard template from the STP storage part 32, and reads
14 out the background noise template from the BNT storage part
15 34. The PF part 36 calculates a residual between the linear
16 combination of templates and the observed profile, its
17 result being registered in the residual storage part 38.
18 Moreover, the sound source localization part 26 specifies
19 the normalized residual giving the minimum residual in the
20 selection part 40 by normalizing the residual stored in the
21 residual storage part 38 and comparing the normalized
22 residuals. Thereafter, the three-dimensional position
23 stored by referring to the standard template giving the
24 corresponding residual is acquired as an appropriate format.

1 Figure 11 is a diagram schematically showing the standard
2 template stored in the STP storage part 32 and the data
3 structure of position coordinates in this invention. The
4 STP storage part 32 is assigned with a memory area
5 corresponding to the three-dimensional position (1, ..., N:
6 N is a positive integer, corresponding to the total number
7 of standard templates). In each memory area i, the STP data
8 and the three-dimensional position data (x, y, z) are stored
9 in association with respective addresses. In another
10 embodiment of the invention, the standard template and the
11 three-dimensional position data may be stored in different
12 memory areas to be referenced from each other.

13 As shown in Figure 11, in the memory area i, the STP data
14 and the three-dimensional position data are stored in
15 association. If the acoustic data is input, the PF part 36
16 converts it into an observed profile, accesses the memory
17 area i in succession to read out the standard template,
18 calculates the linear combination employing the BNT data,
19 and computes the residual between its value and the observed
20 profile, the result being output to the residual storage
21 part 38. In this invention, ~~a delay deformation~~ information
22 defined by the sound reflecting element employed in the
23 invention is introduced into the STP data stored in the STP
24 storage part 32, whereby the elevation is given the

1 intrinsic delay ~~deformation~~ information and acquired with
2 high precision. The selection part 40 refers to the memory
3 area i giving the minimum residual, and reads out the
4 three-dimensional position data (x, y, z) stored in the
5 memory area i to acquire the sound source position. The
6 acquired three-dimensional position data is made a control
7 input into the application execution part 42 to control the
8 driving of the driving element 28, as shown in Figure 11.

9 [Example Embodiments]

10 Specific embodiments of the invention will be described
11 below by way of example, but the invention is not limited to
12 the following examples.

13 (Example 1)

14 Sound reflecting element for acquiring the elevation
15 in the forward direction

16 Assuming that the azimuth of a sound source candidate
17 position was 90° (forward direction), the range to a sound
18 source was 2 m, and the acquirable elevation was from 0° to
19 72°, an enveloping surface of the spheroid was produced as
20 the sound reflecting element. An upper end portion of the

1 sound reflecting element formed in Example 1 reflects a
2 sound wave from the sound source position at high elevation
3 to converge into the microphone position and a portion near
4 the root of the sound reflecting element reflects a sound
5 wave from the sound source position at low elevation to
6 converge into the microphone position. On the other hand,
7 the sound wave from other sound source positions is
8 diffused. If the reflecting position is different, a stroke
9 difference from the direct wave is also varied, generating a
10 proper reflected wave with a delay amount corresponding to
11 the sound source position added.

12 In the case in which the sound reflecting element was
13 employed, there was a delay time difference of about 0.28 ms
14 (milliseconds) in the path difference between the direct
15 wave and the major reflected wave, when the elevation to the
16 sound source was 0° and 72° . The sound source localization
17 system was composed of the sound reflecting element, the
18 microphone, the AD converter, and the microcomputer, whereby
19 the precision of the acquired sound source position was
20 examined. The sampling frequency of the sound source
21 localization system was 48 KHz, and the elevation resolution
22 in which the elevation to the sound source was from 0° to 72°
23 was made discernable at 13 levels at maximum.

1 (Example 2)
2 Confirmation for generating a "delay ~~deformation~~
3 information" in the sound reflecting element

4 The sound reflecting elements formed in Example 1 were
5 disposed as shown in Figure 7, and had two microphones
6 attached to form a sound collecting recording part of the
7 invention. For the input, the voices were used, speakings
8 "there" and "hello" for several seconds were regenerated
9 from the sound source position in the forward direction and
10 with the range 2 m and the elevation 0°, 15° 30°, 45°and 60°,
11 whereby an observed profile was produced as the input voice.
12 At this time, the sampling frequency was 48 KHz. To confirm
13 the existence of reflected wave having delay ~~deformation~~
14 information of the invention, one of the analysis methods of
15 high sensitivity, CSP (Cross-power Spectrum Phase analysis)
16 method by M. Omologo et al. ("Acoustic event localization
17 using a cross power-spectrum phase based technique.", proc.
18 ICASSP 94, pp. 273-276, 1994.) was employed.

19 The CSP method, which traces the acoustic signal at high
20 sensitivity, can give the delay ~~deformation~~ information at
21 high sensitivity in this invention. For the sound source at
22 an elevation of 30°, the calculated CSP coefficients will be
23 shown. Since the CSP method generates a number of pseudo

1 peaks, it is optional how small sub-peak relative to the
2 main peak should be regarded as the valid peak, unlike the
3 main peak. At present, the peaks having one-tenth or more
4 the intensity of the main peak and upper intensities to the
5 third were set as the effective peak.

6 Figure 12 shows the CSP coefficients obtained from the input
7 sound signal for the sound source having an elevation of 30°.
8 The results are shown in Table 1.
9

1 [Table 1]

2 Table 1 Peak positions detected by CSP method (unit: number
3 of samples)

Elevation of sound source→	0°	15°	30°	45°	60°
First place peak position	0	0	0	0	0
Second place peak position	N/A	10	9	6	2
Third place peak position	N/A	N/A	N/A	-6	-
Sub-peak position expected on design	±14	±12	±9	±5.5	±2.5

4
5 The peak position having the first place intensity
6 corresponds to the direct wave, in which the peak position 0
7 indicates that the sound source is disposed in the direct
8 front. At the second place and third place peaks, it is
9 expected that two sub-peaks due to correlation between the
10 direct wave and the reflected wave are detected at the
11 position of designed point as indicated in the table. In
12 Example 2, at least one sub-peak having significant
13 intensity was detected in the cases except for 0° as
14 indicated in the table 1. Also, the delay ~~deformation~~
15 information for the sound source position was detected by
16 detecting the existence of the expected sub-peak to
17 correspond to the designed point. In the case of the sound

1 source elevation of 0° , the expected sub-peak position was
2 not detected. The reason is that the sound reflecting
3 element formed in Example 1 has a reflection area of zero
4 designed for an elevation of 0° (the root of the sound
5 reflecting element).

6 Figure 13 shows a correlation between the sub-peak position
7 obtained in Example 2, and the sub-peak position expected on
8 design. As shown in Figure 13, the observed sub-peak
9 position has the fine correlation with the existing position
10 of the reflected wave expected in the sound reflecting
11 element of Example 1. From the result of Figure 13, it is
12 found that the sound reflecting element formed in Example 1
13 gives an expected delay ~~deformation~~ information.

14 (Example 3)

15 Employing the sound reflecting element formed in Example 1,
16 an examination was made to determine whether or not the
17 elevation of sound source could be practically acquired
18 correctly. For the acquisition of sound source position
19 using the delay ~~deformation~~ information, the PF method was
20 employed in this Example 3. A white noise was regenerated
21 from a noise sound source at a horizontal angle 75° , a range
22 1 m, and an elevation 0° to simulate the background noise.

1 The speaking utterances and the sound levels from five
2 positions were produced by changing the elevation, with the
3 background noise superposed, to create the test voices.
4 Employing the following formula, the score was defined from
5 the view point of what difference is provided for the second
6 best candidate, whereby the precision of acquiring the
7 elevation position was examined. Where n^* is an identifier
8 of the standard template corresponding to the correct
9 position, and the residual Φ_n^* is the normalized residual at
10 the correct position.

11 [Formula 4]

$$12 \quad \rho = \frac{\bar{\Phi}_{\bar{n}} - \bar{\Phi}_{n^*}}{\bar{\Phi}_{\bar{n}}} \quad (4)$$

13 [Formula 5]

$$14 \quad \bar{n} = \underset{n \neq n^*}{\operatorname{argmin}} (\bar{\Phi}_n) \quad (5)$$

15 The above score is given 100% if the normalized residual is
16 zero when the profile corresponding to the correct sound
17 source candidate position is selected, and given 0% or less
18 when the acquisition of sound source candidate position
19 fails, because the normalized residual for another profile
20 has the minimum value.

21 In Example 3, the averaging operation of the sub-band when

1 calculating the normalized residual was made in a range from
2 985 Hz to 7504 Hz where the influence of the sound
3 reflecting element is most apparent. The results obtained
4 are shown in Figure 14. As shown in Figure 14, in any case,
5 one correct sound source candidate position can be selected
6 from among the five candidate positions by exploiting the
7 component decomposition by the PF method, without being
8 affected by the noise. Also, in this invention, when the
9 background noise template is not employed, the score are
10 decreased with the decrease of the S/N ratio. In this
11 invention, the acquisition of sound source position is made
12 with high precision regardless of the S/N ratio by
13 incorporating the background noise template for the residual
14 calculation.

15 Though this invention has been described above by way of
16 example, the invention is not limited to the above described
17 examples. It will be understood to those skilled in the art
18 that various changes and exclusions, and other examples may
19 be made. Also, the sound source acquisition method of the
20 invention can be described in any programming language as
21 ever known, in which these languages include C, C++,
22 Assembler and machine language. Also, the program that can
23 be executed by the computer to perform the sound source
24 acquisition method of the invention may be stored in ROM,
25 EEPROM, flash memory, CD-ROM, DVD, flexible disk, or hard

1 disk and distributed.